

# Automated process of determining “black points” in traffic using business intelligence systems

Atanasijevi Jordan\*, Atanasijevi Nevena\*\*

\* General Staff of the Serbian Armed Forces, GARDA, Belgrade

\*\*ZU AP „Oaza zdravlja“, Beograd

jodzavogi@hotmail.com, nevenastankovicvr@gmail.com

**Abstract - The basic setting of the paper is related to the improvement of the current way of decision-making by using the business intelligence system, important for decision-making on a specific problem, and in order to achieve benefits in decision-making at the highest level. The improvement must be appropriate to the development of the BI system in conditions when there is already previously accumulated knowledge about the problem. The obtained results should show that the business intelligence system can identify dangerous places on the roads, as the first and very important stage in the process of managing "black points", and is a procedure for detecting specific locations on the road network where appropriate concrete measures need to be taken.**

## I. INTRODUCTION

Data warehousing and business intelligence are techniques that provide business people with the information and tools they need to make operational and strategic business decisions.

The beneficiaries are mostly business people in companies. But not all business people carry the same meaning. Special care should be taken for those making strategic business decisions. One good business decision can bring millions of dollars to companies. That is why the main users are executives, managers and analysts in companies. Data warehouse and business intelligence (DW/BI) are high-impact systems.

The term strategic also means importance. These are decisions that can bring profit, but also loss to the company. Therefore, the DW/BI system is a high business risk system. When a strategically important decision is made, someone wins and someone often loses.

The DW/BI system also supports operational decision making, especially when the decision maker has to see a data archive or integrated data from multiple sources. Many analytics applications have such an operational focus. Whether decision-making is strategic or operational, the DW/BI system must provide the information needed to make those decisions.

Decisions usually require a single subset of information, which is basically not predetermined. It is necessary to build an information infrastructure that integrates data from the entire organization and potentially outside the organization, and then cleans, corrects and restructures the data to be as flexible and usable as possible. While most transaction system modules work with one type of information, such as invoices, orders, or receivables, the DW/BI system must integrate them all together. The DW/BI system requires technically sophisticated data management and collection.

Finally, it is necessary to provide the decision maker with the tools to use the data. In this case, the term "tools" is much more than just software. It means in this case everything that users need to be able to understand what information is available to them, then to be able to find a subset of information that they need and finally to be able to structure the data to see the business dynamics. "Tools" include training, documentation and support, along with ad hoc inquiries, reports and analytical applications.

## II. THE GOAL OF THE WORK

The main goal of the paper can be defined as exploring the possibility of identifying and determining "black points", primarily using the software package SQL Server Integration Services (SSIS) which allows automatic extraction, transformation and loading of data, ie development of ETL systems and detection of legality in data numerous analyzes and the use of results in order to make timely and correct decisions and Power BI for data visualization.

A special goal of the paper is the practical implementation of the obtained new and unified approach in order to use the same in order to predict future conditions and achieve optimal operating parameters in different situations when making business decisions

## III. BUSINESS INTELLIGENCE

The term business intelligence was first used in 1996 to denote the category of means of data analysis, querying, reporting, which can help the user in business processes to synthesize valuable information from a huge amount of data on which to base their business decisions.

The phenomenon of business intelligence can be observed from two aspects - macro and micro aspects. Viewed from the macro aspect, business intelligence means a complex aggregate category, which is created by systematic, but not targeted collection of data on macroeconomic trends in a particular geopolitical environment. It also involves their organization and structured recording, as well as logical-computational processing to detect trends. Today, the special attention of knowledge engineers is increasingly aroused by the phenomenon of business intelligence observed from the micro aspect. In this case, it is about discovering hidden knowledge from business data, which an organization collects routinely, performing its daily business transactions.

Business intelligence is a relatively young term for which there are multiple definitions, but each boils down to the fact that business intelligence is a process of collecting data, turning that data into information that is useful for making important decisions. Business intelligence as a term is most commonly used to denote

computer decision support. The business intelligence system is part of the organization's information system specifically designed to support management. Management requires a comprehensive and timely insight into the indicators of the functioning of the organization in order to make decisions reliable and accurate. According to modern theories, this insight should enable as many employees as possible, from whom, when they receive it, greater efficiency and responsibility for the achieved results can be expected. Business intelligence techniques (data warehousing, reporting, OLAP, data mining, dashboards, etc.) extract data from the existing information system and transform it into a form suitable for decision making. The implementation of BI techniques increases the use value of the existing information system of the organization, as a result of which the interest in business intelligence systems is great and still growing.

Since there is no universal definition of business intelligence, different authors define it in different ways. One of the most commonly used and general definitions is the following: "Business intelligence is the use of data that leads to better business decisions. It is about accessing, analyzing and discovering new opportunities."

#### IV. DATA WAREHOUSE ARCHITECTURE MODEL USED

The split requirement plays a key role in defining a typical architecture in a data warehouse system, as shown in the following figure (Figure 1). [1] Although commonly referred to as a two-tier architecture, to emphasize the separation between physically available sources and data warehouses, it actually consists of the following four phases that data goes through:

1. Source layer: The data warehouse system uses heterogeneous data sources. This data is usually stored in corporate relational databases and can come from information systems that are outside the corporate walls.
2. Data transfer: Stored data in sources should be extracted, cleared of inconsistencies, fill in gaps and integrated to merge heterogeneous sources into a common scheme. The so-called ETL (Extraction, Transformation and Loading tools) can merge a heterogeneous scheme, extract, transform, clean, validate, filter and load data from a source into a data warehouse. Technically speaking, this phase deals with problems that are typical of distributed information systems, such as inconsistent data management and incompatible data structures.
3. Data warehouse layer: Information is stored in one logical centralized warehouse, a data warehouse. The data warehouse can be accessed directly, but can also be used as a source to create data marts, which partially replicate the contents of the data warehouse and are designed for specific departments of the enterprise.

4. Analysis: In this layer, integrated data is efficiently and flexibly available for reporting, dynamic information analysis, and simulation of hypothetical business scenarios. Technically speaking, they should have the following features: navigation over aggregated data, complex query optimizers, and a user-oriented GUI.

The architectural difference between data warehouses and data-marts should certainly be clearly expressed. The component marked as the data warehouse in the following figure (Figure 1) is also often referred to as the primary data warehouse or corporate data warehouse.

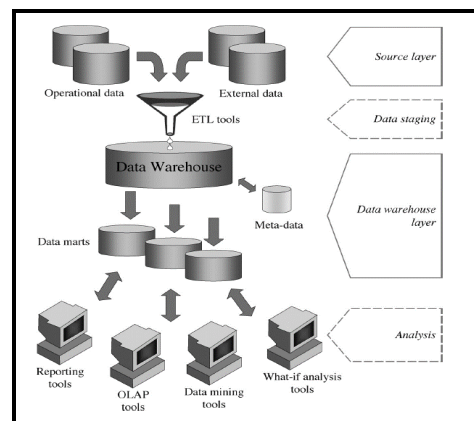


Figure 1. Data warehouse architecture

Any analysis of actual data includes data manipulation, visualization, and modeling. Visualization and modeling are complementary. Visualizations will help us improve unclear issues and rely on human interpretation. The models are much better and allow more complex calculations, but are limited by their assumptions. Software package Microsoft Power BI is very suitable for this type of analysis and presentation of results, in a user-friendly way.

The end product of the analysis is not a model, but rhetoric. The analysis is meaningless if it does not convince someone to take action. In business, this usually means convincing senior executives who have little statistical knowledge, when making a decision.

#### V. PORTAL OF OPEN DATA

The development trend of the Republic of Serbia follows the comparison and monitoring of the best and most advanced countries in all fields, including in traffic. For these reasons, it is necessary to act strategically in order to activate the traffic safety system in the Republic of Serbia. The Republic of Serbia has started to open its data, and the competent institution responsible for the introduction of open data is the Office for Information Technology and Electronic Governance within the Directorate for Electronic Governance.

Open data is digital data available to the public. They have such technical and legal characteristics that anyone,

at any time and anywhere, can use, reuse and redistribute them. Open data can help governments, citizens and organizations achieve better results in the field of public services.

Datasets in the open data lookup have been provided in .CSV format, which is extremely suitable for transformation, compression and analysis, and is also available for conversion to many other formats suitable for data analysis.

The published document on the number of traffic accidents that occurred on the territory of the City of Belgrade in 2018 contains seven variables (columns). The names of the columns are defined in the following table.

Table 1. Names and explanations of variables found in the open data set used in the survey

Ordinal	The name of the variable	Explanation
1.	<b>NEZG_ID</b>	Traffic accident
2.	<b>VREME_NEZ</b>	Date and time of traffic accident
3.	<b>VRSTA_NEZ</b>	Type of traffic accident
4.	<b>NAZIV_TIP</b>	Name and type of traffic accident
5.	<b>NAZIV_DET</b>	A more detailed description of a traffic accident
6.	<b>WGS_X</b>	X coordinates of the WGS (World Geodetic System) or Longitude
7.	<b>WGS_Y</b>	Y coordinate of the WGS (World Geodetic System) or Latitude

**NEZG\_ID** - A traffic accident identification number, which is presented as a seven-digit number.  
**VREME\_NEZ** - The date and time of the traffic accident is given in the format dd.mm.yyyy, hh: mm (day.month.year, hour:minute). From the given data, it can be concluded that the complete data are available in the database for all twelve months of 2018.

**VRSTA\_NEZ** - The type of traffic accident is divided into: 1) traffic accidents with material damage; 2) traffic accidents involving injured persons; 3) traffic accidents with dead persons.

**NAZIV\_TIP** - The name and type of traffic accident is divided into: 1) single-vehicle traffic accident; 2) traffic accidents with at least two vehicles (no turning); 3) traffic accidents with at least two vehicles (turning or crossing); 4) traffic accidents with parked vehicles; 5) pedestrian accidents.

**NAZIV\_DET** - A more detailed description of a traffic accident contains 68 events: 1) a single-vehicle traffic accident (11 events); 2) traffic accidents with at least two vehicles (no turning) (9 events); 3) traffic accidents with at least two vehicles (turning or crossing) (18 events); 4) traffic accidents with parked vehicles (5

events); 5) pedestrian accidents (25 events).

## VI. RESEARCH RESULTS

SQL Server Integration Services (SSIS) was used for the development of the data warehouse, which enables an automated process of data extraction, transformation and loading, ie the development of the ETL system. As previously mentioned, excel files available on the open data portal were used as data sources.

The following figure shows the process of loading traffic accident data, which will be used later for analysis.

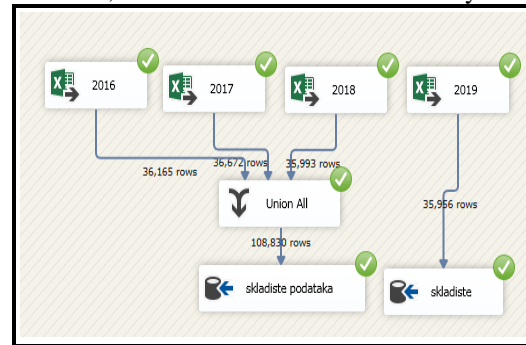


Figure 2. Loading data into the data warehouse

The data used was previously loaded, transformed and stored in SQL Server 2016 DBMS. The flow of data processing is shown in Figure 3.

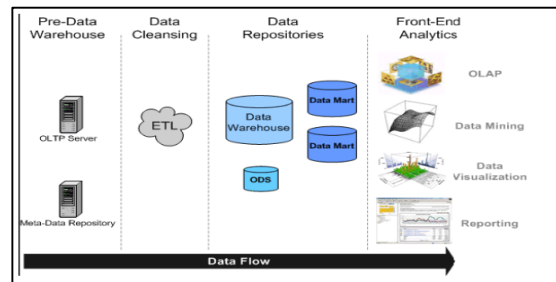


Figure 3. Flow of data processing and data visualization

Dimensional modeling is a logical design technique aimed at presenting data in a standard form that ensures high system performance. The result of dimensional modeling is a dimensional data model that includes the definition of dimensions, hierarchies and relationships. In a dimensional model, data is organized to describe dimensions and dimensions.

Dimensions define each transaction and are stored in tables that are linked to a fact table. Measures are numerical data stored in a fact table. In this way, a data model is created that uses multidimensional arrays to get a view of the data as seen by executives, managers, analysts or planners. Users typically view and analyze data by time period, person, or some similar dimension.

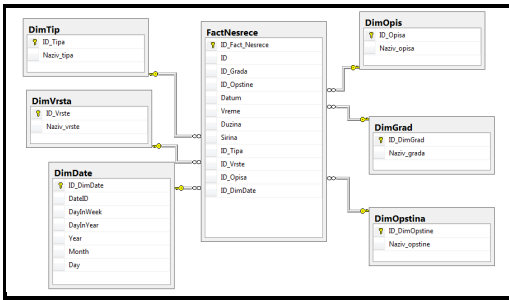


Figure 4. Data warehouse data model

The star scheme stores all dimension information in a single table. Each level of the hierarchy is represented by a column or set of columns in the dimension table. A dimensional object can be used to define a hierarchical relationship between two columns (or a set of columns). Without dimensional objects, hierarchical relationships are defined only in metadata. Attributes are located in the columns of dimension tables. Dimensional tables contain descriptive textual information. Dimensional table attributes are used as constraints when querying. Each dimensional table has its own primary key, and they all participate in the creation of the primary key of the fact table.

Fact tables are used to store measures and contain data that are usually of a numerical type and can contain a large number of records. Fact tables contain a composite primary key, which consists of several foreign keys (one for each dimension table) and columns for each measure that uses these dimensions.

The first step in implementing a dimension is to break the link between the fact table and the dimensions. This is done due to the rules of referential integrity, because otherwise the database management system would not allow deleting data from the table that represents the dimension. After breaking the link between the tables, the table representing the dimension is deleted and re-created, with the same columns, attributes, data types and constraints. In this way, all previous data is deleted and the table is ready for entering the most up-to-date status from the source data.

After creating all the dimensions and the table of facts, we start creating an automated process that unites all the actions. The appearance of one such process is shown in the following figure.

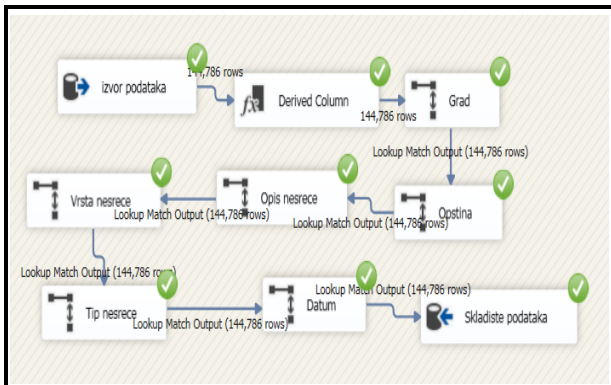


Figure 5. Automated data warehouse filling process

After the data warehouse is filled with the most up-to-date data, the process of automatic loading, transformation and entry of data, the process of determining "black points" is started using SQL Server Management Studio. Cursors were used to determine "black points". Cursors are used to work with individual records. A cursor is a set of records that are formed on a server and associated with a pointer that points to the current row. In this particular case, with the current record representing the location of the traffic accident that occurred, compare and record those traffic accidents that occurred on the road at least 500m, and that there were at least two traffic accidents with fatalities or injuries. The following figure shows the "black dot" around which most traffic accidents occur.

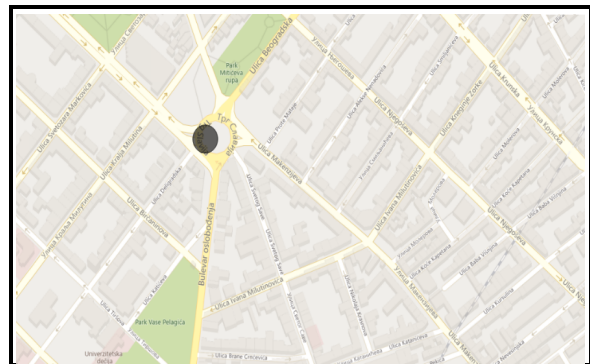


Figure 6. Graphic representation of the "black spot" with the most traffic accidents

Figure 7 shows the appearance of the graphic representation of 3 "black points" on the territory of the city of Belgrade.

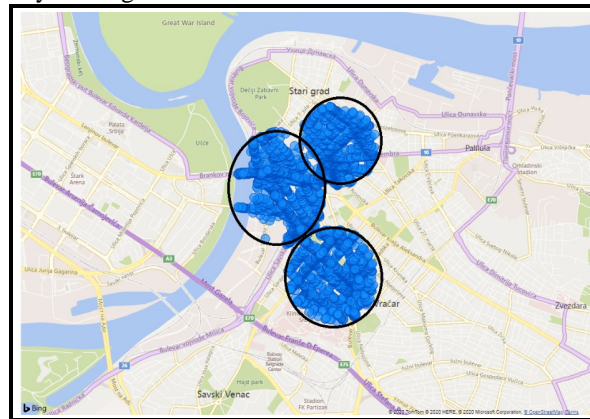


Figure 7. Graphic representation of 3 "black points" on the territory of the city of Belgrade

Figure 8 shows the appearance of a graphic representation of 9 "black points" on the territory of the city of Belgrade.

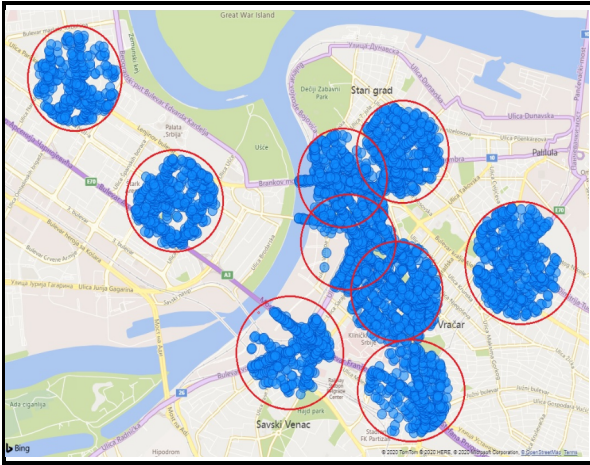


Figure 8. Graphic representation of 9 "black points" in the city of Belgrade

Figure 9 shows the appearance of the graphic representation of 11 "black points" on the territory of the city of Belgrade.

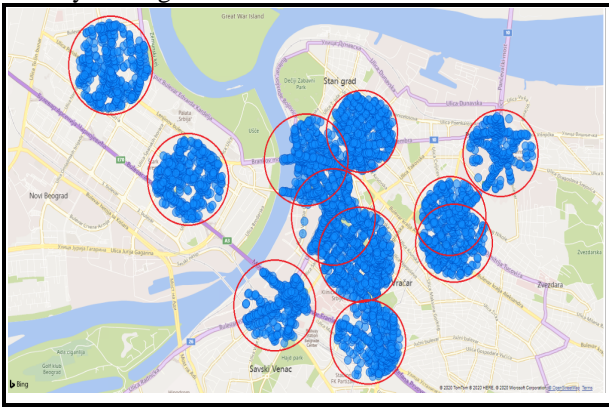


Figure 9. Graphic representation of 11 "black points" in the city of Belgrade

## VII. CONCLUSION

The paper analyzes the current situation in the field of research, using published papers in the previous period that are focused in whole or in part on the elements of using the software package Microsoft Power BI.

When designing and developing a business intelligence system, it is analyzed whether it facilitates the work of end users, whether the data used is expensive and can be easily accessed, whether the information it generates is expensive, whether the data is easily accessible, etc.

At the same time, we must not neglect what is actually necessary for the end user. The end user wants to use "fresh" information, at any time, from any location, and all this with minimal effort to integrate into the existing system. Also, the end user wants a unique and integrated data source, to create reports independently and to have a simple decision support tool.

One of the main causes of insufficient success in the introduction of the business intelligence system is the human factor, ie. the ability of end users to change and adapt to the changes brought about by new technology.

Although there are numerous methods of planning, analysis, design, construction, use and maintenance of business intelligence systems, they do not yet give positive results. It is necessary to find ways and methods in which to change the established way of working, and from keeping records and writing reports in a formalized form (notebooks, reports, ...) to move to keeping data in electronic form.

In order to solve this problem, positive attitudes of people from the management structures of the system are necessary, as well as the acceptance of new systems by end users and professionalism, ie. knowledge of that technology and the possibility of its application. The paper proposes the application of the software package Microsoft Power BI in order to improve the business intelligence system, when the data are previously loaded, transformed and stored in one of the DBMS.

## REFERENCES

- [1] Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2008). *The data warehouse lifecycle toolkit*. John Wiley & Sons.
- [2] Republic of Serbia Open data portal: <https://data.gov.rs/sr/datasets/>, accessed 09th December 2019
- [3] Fotache, M. (2016). Data processing languages for business intelligence. *SQL vs. R. Informatica Eco- nomica*, 20(1), 48.
- [4] Jatain, A., & Ranja, A. (2017). A Review Study on Big Data Analysis Using R Studio. *International Journal of Computer Science and Mobile Computing*, 6(6), 8-13.
- [5] Joseph Rickert, "Big Data Analysis with Revolution R Enterprise", 2011
- [6] Wang, J., & Gu, L. (2016). Challenges of teaching data science in a business school. *Issues in Information Systems*, 17(3).
- [7] Fotache, M., & Strimbei, C. (2015). SQL and data analysis. Some implications for data analysis and higher education. *Procedia Economics and Finance*, 20, 243-251.
- [8] Wickham, H. (2019). Data science: how is it different to statistics?. *IMS Bulletin*, 48.
- [9] Kimball, R., & Ross, M. (2016). *Dimension Table Core Concepts*. The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence, 327-384.